DOCUMENT RESUME

ED 050 156 TM 000 548

AUTHOR Bradley, Paul A.; Woolley, Dale
TITLE Making Better Decisions on Assessed Needs: Differentiated School Norms.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
PUB DATE Feb 71
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Academic Performance, Comparative Analysis, Curriculum Evaluation, Educational Improvement, *Educational Needs, *Evaluation Methods, Evaluation Needs, *Norms, *Program Evaluation, School Improvement, Standardized Tests, Statistical Data

ABSTRACT

This paper discusses the use of student performance data as an input to the last phase of a needs-assessment evaluation. The weaknesses for this purpose of national pupil norms are examined and it is suggested that improved data is required to reflect the many differences that exist between schools. Specifically, two changes are proposed: the provision of national school norms with standardized tests as well as pupil norms, and, if feasible, norms for different "types" of schools. Hypothetical examples are presented in graph form and the uses of such data are discussed. (AE)

Making Better Decisions on Assessed Needs: Differentiated School Norms

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Paul A. Bradley and Dale Woolley

Center for the Study of Evaluation
University of California at Los Angeles

ED050156

The procedures in a needs-assessment evaluation that have been described heretofore include the selection of educational goal areas that are to be evaluated and the selection of the best available instruments to assess student performance in these goal areas. Following the selection of assessment instruments is the administration and scoring of the instruments. The information provided by the assessment devices becomes one of the inputs to the final phase of a needs-assessment evaluation: the selection of the one or more educational goal areas in which revisions in the instructional program will be made so as to improve student performance.

This last phase in the needs-assessment evaluation is the critical one, obviously, as it pinpoints where a school is going to devote some time, effort, and, probably, money to correct a deficiency in its instructional program. Making a bad decision at this phase would have dire consequences, as the expenditures of time, effort, and money in behalf of the selected goal area would be for naught, and another goal area which would have better deserved the attention would have been neglected. Therefore, it is imperative that a school have the best information possible before it decides which educational goal area (areas) is (are) selected as a target area for improving student performance.

One type of information that is an input to the last phase of a needs assessment evaluation is the data from the assessment of student performance. This paper is concerned with ways in which this information can be improved so that it is maximally useful to the school which is involved in a needs assessment evaluation.

TM 000 548

What is the outcome when a school assesses student performance with standardized instruments? First of all, the outcome depends on what is made available by the publisher of the instrument. Since all, or nearly all, publishers provide tables of norms, a school could prepare a roster of the raw and scaled scores achieved by every pupil. It is to be understood from this point on that we are talking about student performance within different grade levels. At no time are we looking at or comparing the performance of students in different grades. This notion follows the common practice of interpreting test results relative to the grade level of the student. The most frequently reported scaled scores are centiles, grade equivalents, and stanines. An ambitious person could then take this roster and compute averages for each grade level, and if this information were available for other schools in the district then he could compare averages across schools. If a battery of tests had been administered then it would be possible to prepare a profile of achievement for every pupil. However, it is uncommon that there is a person in an elementary school who either has the time or the experience to undertake such an endeavor.

The larger publishing houses make available to a school several services that aid in the interpretation of test results. The services include providing information similar to that described above: that is, rosters of scaled scores, various descriptive statistics for grades, schools, or systems, and individual pupil profiles. In addition, a publisher may indicate the procedures to be followed if a school wished to develop percentile scores for students either within a school building or, within a school system. However, most of the information that can be provided by test publishers is useful only for evaluating the current status of individual pupils. That

is, the various scaled scores that publishers provide indicate the goodness of a student's performance relative to the performance of all students who took the same test.

There are two reasons why the information that is most typically available neither is the best information available nor is maximally useful to a school that is engaged in a needs assessment evaluation. One reason is that virtually all currently available test norms are pupil norms, that is, they indicate the relative goodness of an individual student's raw score. The second reason is that, again, virtually all test norms are based on samples of students that are intended to be representative of all students in the country--national norms. Why is it, however, that these reasons make the typical test norms inappropriate for a needs assessment evaluation?

School Norms

In a needs assessment evaluation, the unit being evaluated is the school, not a single student. One aspect of a needs assessment evaluation is determining how good a job a school is doing in producing appropriate student achievement in the chosen educational goal areas. That is, once a school has identified the educational objectives that are most important it must determine its level of achievement in those educational objectives. It is not possible to determine the school's level of achievement from pupil norms, as these norms are inappropriate. What is needed instead, for the purpose of a needs assessment evaluation, are norms that would give the relative goodness of a school's performance on a standardized test. Percentile norms could be derived for schools very easily since the process would be the same as that used in deriving pupil norms. The one change is that it is the schools' mean raw scores on the test rather than the pupils' raw scores that are used to compute percentiles.

One might, at this point, wonder why a school cannot determine its level of performance by looking up its mean raw score on a standardized test in a table of pupil percentile norms. It is not appropriate to do this because a school would get an incorrect indication of its level of performance. The difference between pupil and school norms is based on the fact that there is less variation in school means than in pupil raw scores. Figures 1 and 2 illustrate this difference. Figure 1 shows hypothetical normal frequency distributions of pupil raw scores (A) and school scores (B). It is seen that there is less variation in the school scores than in the pupil raw scores.

A normal frequency distribution was chosen for convenience only. No implication is intended that actual pupil or school frequency distributions have the characteristics of a normal distribution. It is also a convenience that the means of the distributions are the same. The standard deviation of the pupil scores is 10 while that of the school scores is 5. No generalization is possible regarding the ratio of standard deviations of pupil scores and school scores, other than that the former is larger than the latter. Again, the standard deviations of 10 and 5 were chosen for their graphical and conceptual impact.

Figure 2 shows the cumulative proportions of the frequency distributions in Figure 1. The curves in this figure can be used to read the pupil and school percentile scores. For example, if a pupil's score was 24, then his percentile score is 27. But, if a school's score was also 24, then its percentile score is 11. If one looks at a score of 37, however, it is seen that the pupil percentile is 75 and the school percentile is 92. Thus, looking at a raw score that would fall below the 50th percentile, a school's percentile

score is lower than a pupil's percentile score, but looking at a raw score that would fall above the 50th percentile, a school's percentile score is higher than a pupil's percentile score.

Differentiated Norms

There is yet another way in which the norm tables provided by most publishers can be inadequate for a needs assessment evaluation. In most instances the norm tables are based on a national sample of schools, for better or worse. Thus, even if a publisher did produce school percentile norms, a school's performance would be compared to the performance of all schools in the sample. But, what if there were certain characteristics of schools, characteristics outside the student's cognitive and affective skills, that were found to be related to their level of performance? The characteristics of a school that are pertinent here are often referred to as input variables. An example of an output variable would be school performance on a standardized achievement test. The input variables could be such things as the number of volumes in the school library, the average expenditure per student, the occupational level of parents, or the racial mixture of the students. Indeed, it is not even necessary to hypothesize about this situation, because the Coleman study, for example, has shown that such relationships do exist. Under these conditions, then, the use of national norms can lead to an unfair and biased comparision if a school is atypical in its characteristics.

The bias can lead to both over- and under-estimation of a school's level of performance. This bias can be easily illustrated through the use of cumulative proportion curves. Suppose it were found that there were three different "types" of schools which had markedly different performance on a standardized test. Figure 3 shows the cumulative proportion curves (A, B, C) for the three

hypothetical types of schools as well as a cumulative proportion curve for all the schools (D). The three cumulative proportion curves A, B, and C correspond to normal frequency distributions with means of 20, 30, and 40, respectively. All have a standard deviation of 4. The cumulative proportion curve D corresponds to a normal frequency distribution with mean 30 and standard deviation 8. Again, these means and standard deviations were selected for convenience and graphical impact. No implication is intended that these curves represent distinct possibilities. It is not known how much separation of schools can be achieved, how school scores are distributed, and what the relationship among standard deviations is.

Consider the four outcomes indicated in Figure 3. First of all, suppose a school's score was 14. If that school found its percentile rank from curve D on Figure 3, it would find that its performance fell at the 2nd percentile. It is possible that a principal confronted with this result would immediately begin a litany of alibis which account for his school's low performance. Many of these alibis are likely to refer to the inputs into the school, such as a low SES level, a variable ethnic composition, poor tax support, low teacher salaries, etc. However, suppose that curve A in Figure 3 gives the percentile ranks for schools that are similar to the principal's school in terms of input variables. Using this curve (A), the principal would find that his school's performance fell at the 6th percentile. That is, when compared to schools that have similar resources, his school did better than only 6% of the schools. This result should indicate, rather unequivocally, to the principal that his school is not doing a good job in producing student performance in the area that the test measures.

Now consider a school whose score was 21. With reference to curve D, which represents a table of national norms, a score of 21 corresponds to a

percentile rank of 13. Again, it is likely that the principal of this school would be echoing the sentiments of the principal whose school's performance fell at the 2nd percentile. However, when this principal compares his school's score of 21 with similar schools, he learns that his level of performance falls at the 60th percentile, not the 13th percentile. Needless to say, this principal is not likely to be disappointed with such an outcome, and may even be mildly pleased to outrank 60% of the schools.

The third outcome indicated in Figure 3 is a score of 38. The percentile rank of this score is 84, with reference to curve D (national norms). A principal who finds out that his school falls at the 84th percentile may possibly engage in some strutting and issue proclamations attesting to the superiority of his school. Suppose, however, that this school is of a type that is characterized by favorable values on the input variables that are related to student performance. For example, some of the characteristics of this type might be high SES level, high teacher salaries, and an all white student body. The percentile ranks for this type of school are given by curve C in Figure 3. Using this curve, a score of 38 corresponds to a percentile rank of 31. A markedly different picture of this school now emerges. Rather than doing a good job with the students, the school appears to be somewhat deficient in producing student output commensurate with its input characteristics.

Lastly, consider a school whose score is 46. On the national norms this school falls at the 97th percentile, and the principal of this school would probably be jubilant. Being aware of the quality of the input characteristics of his schools, he may wonder if his school's performance is really as good as it seems. This principal, at least, is in a good position, because his score of 46 falls at the 93rd percentile with reference to curve C.

What are the consequences for needs assessment evaluation when a school uses national norms rather than differentiated school norms? In virtually all cases a school would have an incorrect indication of its performance level. In many of these cases, though, a school would reach the same decision to select a particular goal area for curriculum revision no matter what norms were used. But in some of these instances, using the national norms rather than differentiated school norms will lead to two types of errors: selecting a goal area for curriculum revision that in fact does not need it, and not selecting a goal area that in fact does need curriculum revision. It is impossible to predict when these errors will occur because there are inputs other than test scores into the decision making process of selecting a goal area for curriculum revision. (This phase of needs assessment evaluation will be discussed by J. S. Dyer and G. P. Strickland in the following paper.)

It should be reiterated at this point that the notion of having different norms for different types of schools is a notion whose feasibility depends on finding types of schools that differ in their performance on standardized tests. This latter point is important, because while it may be possible to group schools into a small number of categories based on similarities of input characteristics, it may not be the case that there are significant differences in level of student performance. If there is no difference between the groups in level of performance, then there is no need to have three separate norm tables which are essentially identical to each other. It should be remembered, though, that the situation is different with regard to the notion of having tables of school norms as well as tables of pupil norms. This notion is not only very feasible and plausible, it has been done by some publishers.

To summarize, then, we think that improvements need to be made in the information that results from an assessment of student performance. It is important to improve this information because it is an input to the last phase of a needs assessment evaluation: the selection of the one or more educational goal areas in which revision in the instructional program will be made so as to improve student performance. Specifically, it was proposed that improvements can be made by altering the types of norms that accompany standardized tests. The two alterations suggested were to provide school norms as well as pupil norms and to provide, if feasible, norms for different "types" of schools as well as national norms.

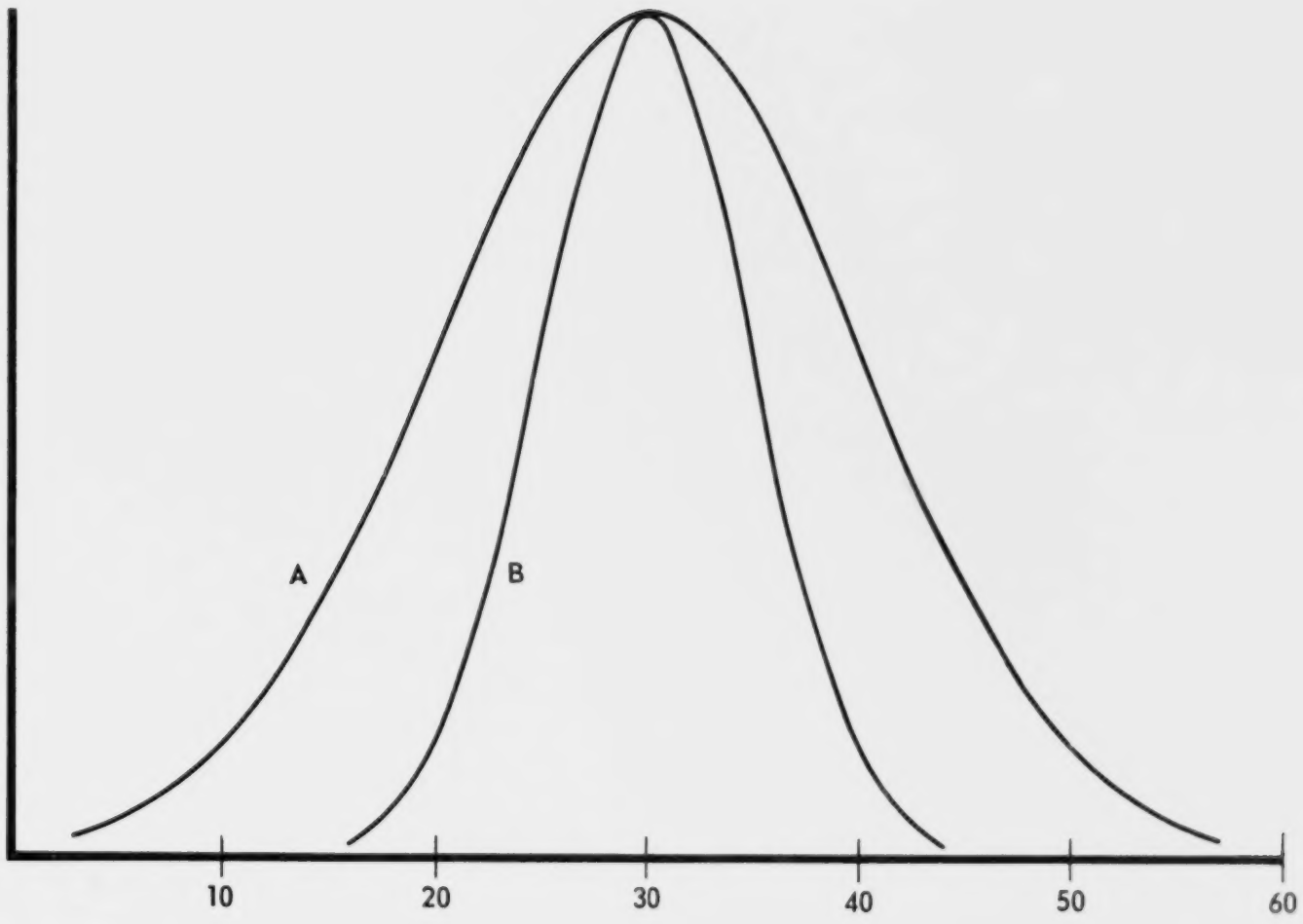

FIGURE 1. HYPOTHETICAL FREQUENCY DISTRIBUTIONS OF PUPIL SCORES (A) AND SCHOOL SCORES (B).

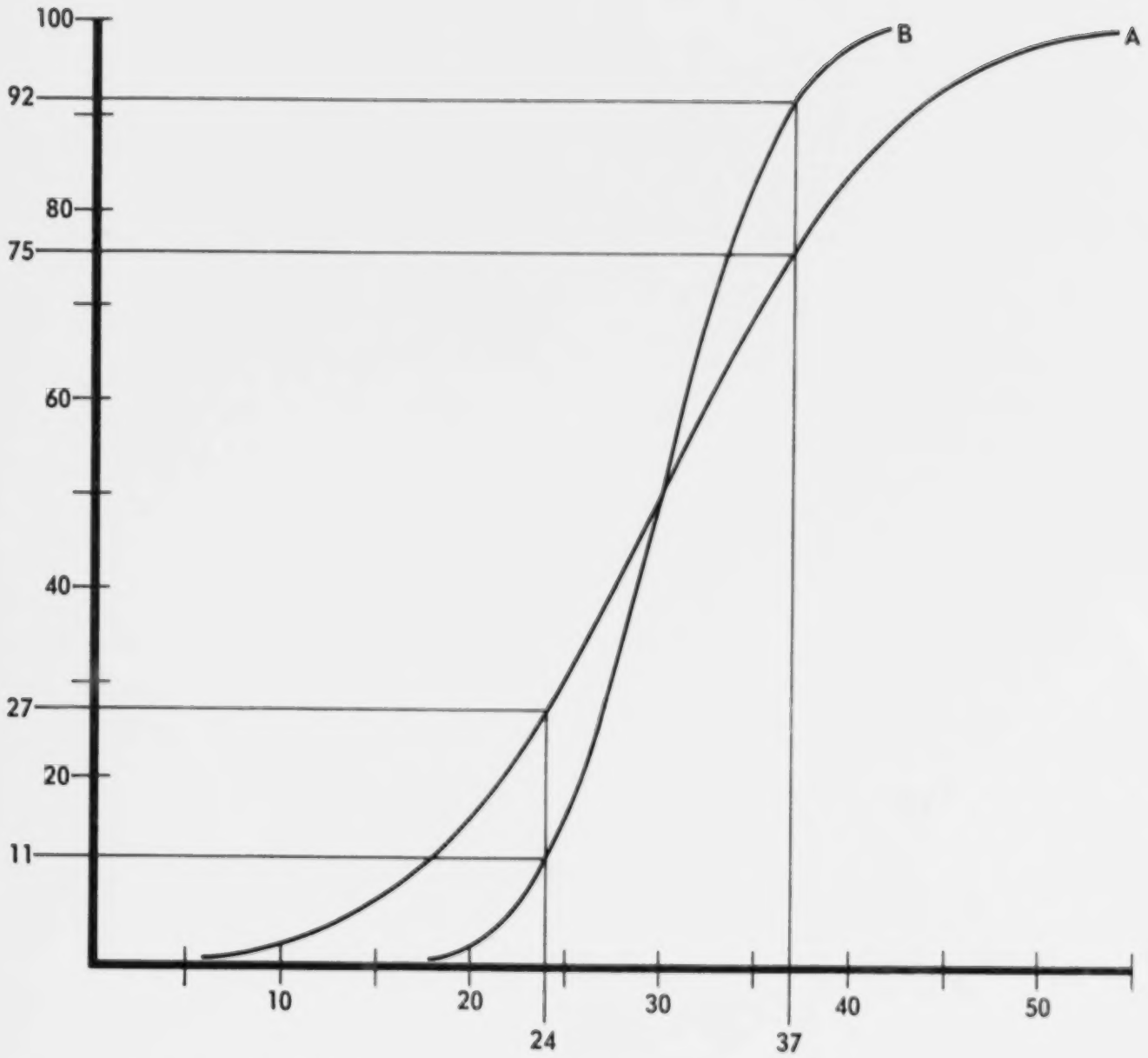

FIGURE 2. HYPOTHETICAL CUMULATIVE PROPORTIONS OF PUPIL SCORES (A) AND SCHOOL SCORES (B).

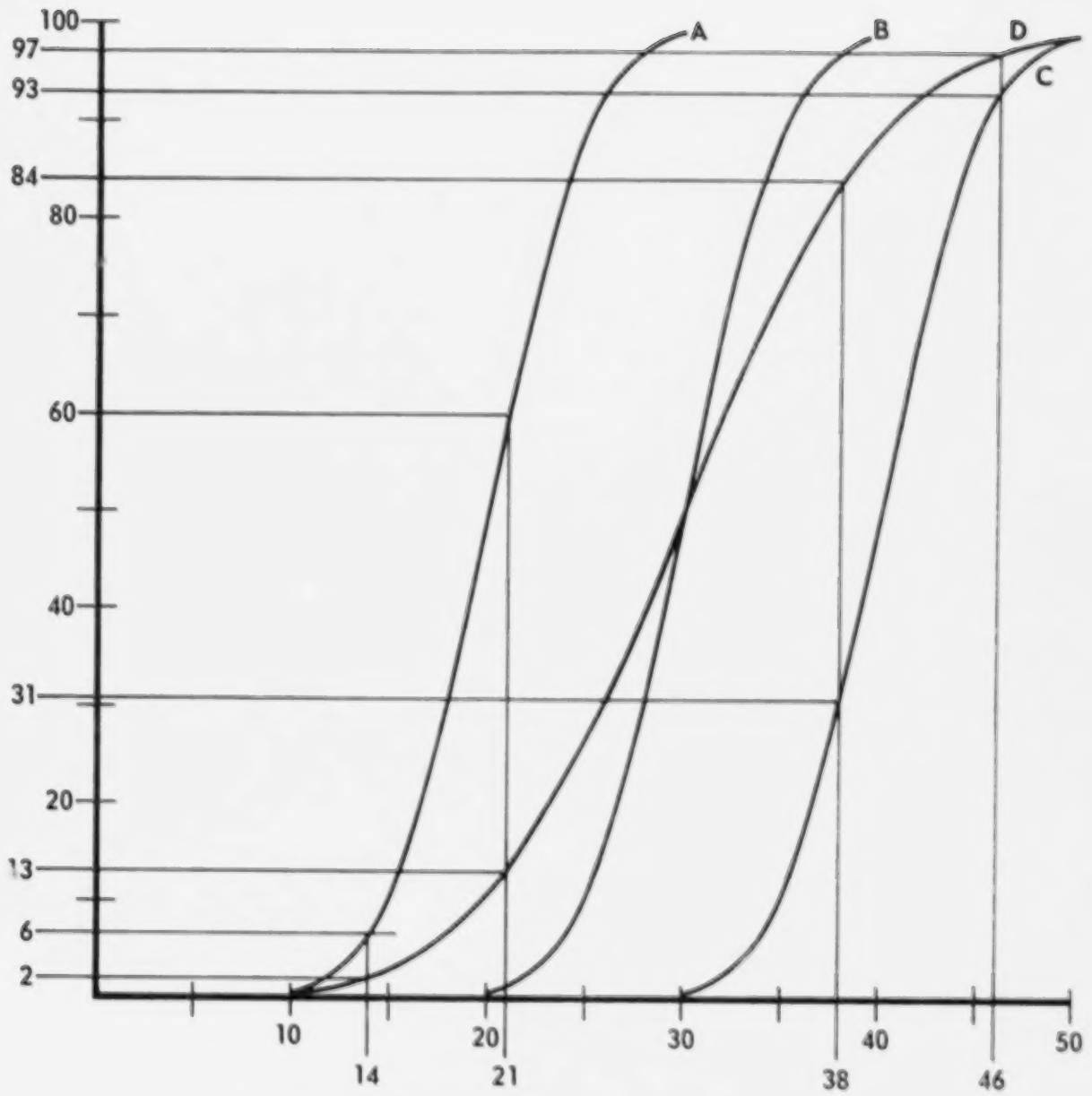

FIGURE 3. HYPOTHETICAL CUMULATIVE PROPORTIONS OF SCHOOL SCORES FOR THREE DIFFERENT TYPES OF SCHOOLS (A, B, C) AND FOR ALL SCHOOLS (D).